Final Project Submission and Evaluation - Dec 12, 2024

The final project submission and presentation are due on December 12, 2024. Each group will have 15 minutes to present their work. During the presentation, you should emphasize the following:

1. Comparison of the technology stack, including models, cloud services, GPU, RAM environment, fine-tuning parameters, dataset, and so on, between your midterm and final submissions.

2. Fine-tuning details, such as epochs, hours, and evaluation results.

3. Demonstrating inference on a set of prompts related to your chosen task using:
- Vanilla QLoRA model
- Fine-tuned QLoRA model
- Vanilla Large Model
- Fine-tuned Large Model

NOTE: The demo must be a live demo in front of the class - a pre-recorded video demo will not be graded.

4. Producing, analyzing, and presenting statistical results for at least 50 prompts and their inferences on the Full Model.

Final Project Submission and Presentation Rubric (40 Points Total)

1. Technical Comparison (10 Points)

- **Comprehensive comparison of tech stack (4 points):**
  Includes discussion on models, cloud services, GPU, RAM, fine-tuning parameters, dataset changes, etc.
  - 4: Detailed comparison with clear contrasts and insights.
  - 3: Adequate comparison, but lacks depth or some key aspects.
  - 2: Superficial or incomplete comparison.
  - 0-1: Minimal or no effort to compare.
- **Link to midterm work (3 points):**
  Highlights evolution from mid-term submission to final submission.
  - 3: Clearly demonstrates how changes improved the project.
  - 2: Some mention of changes but lacks clarity or detail.
  - 1: Barely connects midterm and final work.
  - 0: No connection made.
- **Clarity and structure (3 points):**
  - 3: Comparison is well-organized and presented in an understandable way.
  - 2: Mostly clear but has minor gaps or disorganization.
  - 1: Difficult to follow or poorly structured.

- 0: Not presented.

2. Fine-Tuning Details (10 Points)

- **Explanation of fine-tuning process (5 points):**
  Includes epochs, hours, and evaluation results.
  - 5: Complete and well-detailed explanation with justification for choices.
  - 3-4: Adequate explanation but missing some details.
  - 1-2: Superficial explanation with major gaps.
  - 0: Not explained.
- **Evaluation of results (5 points):**
  Analysis of fine-tuned model performance.
  - 5: Clear and thorough analysis with meaningful insights.
  - 3-4: Partial analysis with some insights but lacking depth.
  - 1-2: Limited or superficial analysis.
  - 0: No analysis provided.

3. Live Demonstration (10 Points)

- **Vanilla and fine-tuned Large Model inference (5 points):**
  - 5: All inferences completed and demonstrated live without issues.
  - 3-4: Most inferences demonstrated live with minor issues.
  - 1-2: Limited or poorly conducted live demonstration.
  - 0: No live demo presented.
- **Effectiveness of live demo (5 points):**
  - 5: Smooth, engaging, and well-structured demo with clear explanations.
  - 3-4: Demo mostly effective but has minor gaps.
  - 1-2: Demo difficult to follow or poorly executed.
  - 0: Demo fails or is missing.

4. Prompt Analysis and Results (10 Points)

- **Statistical analysis of 50 prompts (5 points):**
  - 5: Thorough and insightful analysis with visualizations and interpretation.
  - 3-4: Adequate analysis with basic insights but lacks depth.
  - 1-2: Superficial or incomplete analysis.
  - 0: No analysis conducted.
- **Presentation of results (5 points):**
  - 5: Clear, professional, and engaging presentation of statistical results.
  - 3-4: Good presentation but has some gaps in clarity or professionalism.
  - 1-2: Poorly presented or disorganized results.
  - 0: Results not presented.